

## ORIGINAL ARTICLE

# Global clinical response in Cushing's syndrome patients treated with mifepristone

Laurence Katznelson\*, D. Lynn Loriaux†, David Feldman\*, Glenn D. Braunstein‡, David E. Schteingart§ and Coleman Gross¶

\*Medicine and Neurosurgery, Stanford University, Stanford, CA, †Medicine, Oregon Health Science University, Portland, OR, ‡Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, §Medicine, University of Michigan, Ann Arbor, MI and ¶Concept Therapeutics Menlo Park, CA, USA

## Summary

**Objective** Mifepristone, a glucocorticoid receptor antagonist, improves clinical status in patients with Cushing's syndrome (CS). We examined the pattern, reliability and correlates of global clinical response (GCR) assessments during a 6-month clinical trial of mifepristone in CS.

**Design** *Post hoc* analysis of secondary end-point data from a 24-week multicentre, open-label trial of mifepristone (300–1200 mg daily) in CS. Intraclass correlation coefficient (ICC) was used to examine rater concordance, and drivers of clinical improvement were determined by multivariate regression analysis.

**Patients** Forty-six adult patients with refractory CS along with diabetes mellitus type 2 or impaired glucose tolerance, and/or a diagnosis of hypertension.

**Measurements** Global clinical assessment made by three independent reviewers using a three-point ordinal scale (+1 = improvement; 0 = no change; –1 = worsening) based on eight broad clinical categories including glucose control, lipids, blood pressure, body composition, clinical appearance, strength, psychiatric/cognitive symptoms and quality of life at Weeks 6, 10, 16, and 24.

**Results** Positive GCR increased progressively over time with 88% of patients having improved at Week 24 ( $P < 0.001$ ). The full concordance among reviewers occurred in 76.6% of evaluations resulting in an ICC of 0.652 ( $P < 0.001$ ). Changes in body weight ( $P < 0.0001$ ), diastolic blood pressure ( $P < 0.0001$ ), two-hour postoral glucose challenge glucose concentration ( $P = 0.0003$ ), and Cushingoid appearance ( $P = 0.022$ ) were strong correlates of GCR.

**Conclusions** Mifepristone treatment for CS results in progressive clinical improvement. Overall agreement among clinical reviewers was substantial and determinants of positive GCR

included change in weight, blood pressure, glucose levels and appearance.

(Received 18 April 2013; returned for revision 19 July 2013; finally revised 21 August 2013; accepted 12 September 2013)

## Introduction

Cushing's syndrome (CS) is a serious multisystem disorder due to prolonged exposure to elevated levels of cortisol. Patients with CS experience characteristic signs and symptoms, which often include metabolic disturbances (type 2 diabetes mellitus, impaired glucose tolerance, insulin resistance, dyslipidaemia, obesity), cardiovascular complications (hypertension, hypercoagulability), neuropsychiatric abnormalities (depression, cognitive deficit, psychosis), reduced quality of life, altered physical appearance (moon facies, striae, acne, hirsutism), osteoporosis and reproductive abnormalities (hypogonadism, amenorrhoea). Surgical cure of the underlying aetiology of CS (tumoural source of adrenocorticotrophic hormone [ACTH] or adrenal tumour) generally improves these abnormalities; however, residual hypercortisolism is present in up to half of patients after surgery,<sup>1–3</sup> necessitating the use of adjuvant medical therapy in many patients. Similar to determining success of surgery in CS, resolution of hypercortisolism has been the main measure of response to medical therapy because it is widely believed that normalizing cortisol levels will resolve the signs and symptoms of the disorder.<sup>4–7</sup>

Mifepristone, a glucocorticoid receptor antagonist, blocks the actions of cortisol at the tissue level without a concomitant reduction in serum cortisol. In the recent SEISMIC (Study of the Efficacy and Safety of Mifepristone in the Treatment of Endogenous Cushing's Syndrome) study, administration of mifepristone to patients with CS resulted in improvement in dysglycaemia and hypertension.<sup>8</sup> This study also showed that the global clinical response (GCR), as determined by an independent data review board that assessed overall clinical response based on data from eight clinical categories, including glucose homeostasis, blood pressure, lipids, body composition, strength,

Correspondence: Laurence Katznelson, 875 Blake Wilbur Drive, Stanford, CA 94305-5821, USA; Tel.: 650 721 1020; Fax: 650 736 8100; E-mail: lkatznelson@stanford.edu

psychiatric and cognitive function, physical appearance, and quality of life, improved in 87% of patients. We now report further analysis of the GCR in SEISMIC, including stratification based on clinical outcomes, time dependency and correlations with clinical improvement during mifepristone administration.

## Methods

### Patients and study design

The study was approved by the Institutional Review Board at each of the 17 research sites and conducted in accordance with the Declaration of Helsinki. Fifty patients with ACTH-dependent and ACTH-independent CS were enrolled in a 24-week open-label study of mifepristone, as previously described.<sup>8</sup> Inclusion criteria included active hypercortisolism (defined as an elevated 24-h urinary free cortisol excretion on two collections) with clinical signs and symptoms of CS, including type 2 diabetes mellitus (T2DM) or impaired glucose tolerance (IGT), and/or a diagnosis of hypertension (HT), and the presence of two of six of the following signs and symptoms: Cushingoid appearance (moon facies, dorsal cervical fat pad, facial plethora), increased body weight, proximal muscle weakness, low bone mineral density, psychiatric symptoms, hirsutism and/or pigmented striae and/or acne. For the purposes of the analysis of efficacy, there were two primary end-points that were applied to patients based on the presence of CS and T2DM/IGT (the C-DM cohort) or a diagnosis of CS and hypertension without T2DM/IGT (the C-HT cohort).

Mifepristone was initiated at 300 mg once daily orally and increased in increments of 300 mg daily at Day 14, Week 6 and Week 10 to a maximum of 1200 mg once daily. Investigators could stop escalation based on tolerability and/or if the patient had achieved sufficient clinical benefit at a given dose. The dose could be decreased to manage adverse events (AEs) as previously described.<sup>8</sup> End-points were measured at baseline, Week 6, Week 10, Week 16 and Week 24. These assessments included items from eight broad clinical categories, as shown in Table 1. Investigators graded Cushingoid appearance (facial plethora, moon-shaped facies, dorsal cervical fat pad) on a 0–3 point scale: 0 = absent, 1 = mild, 2 = moderate, 3 = severe at each visit. The interval change was assessed on a five-point scale: much improved, somewhat improved, no change, somewhat worse and much worse. The other end-point measures have been previously described.<sup>8</sup> An independent, three-member data review board (G.B, D.L.L., D.F.) reviewed the clinical data and assessed the overall clinical response. Each board member independently reviewed the data on each visit and compared those findings to the baseline visit; the visits were presented in random sequence blinded to dose and duration of treatment. Raters scored each postbaseline visit for overall clinical status on a three-point ordinal scale: +1 = clinically significant improvement; 0 = no change; -1 = worse from baseline. The median score of the three raters was used to determine GCR: a median score of +1 at any visit was defined as a positive response. Dual-energy X-ray absorptiometry (DXA) data for body composition and

**Table 1.** Clinical parameters assessed by raters for global clinical response

Glucose
Fasting plasma glucose test
Oral glucose tolerance tests*
Haemoglobin A1c test†
Change in glucose-lowering medications
Lipids
Total cholesterol and lipoprotein fractions and triglycerides
Blood pressure
Systolic blood pressure
Diastolic blood pressure
Change in antihypertensive medications
Body composition/bone
Body weight
Body mass index
Waist circumference
Bone resorption and formation markers
Clinical scores/appearance
Investigator-graded Cushingoid appearance‡
Investigator-graded acne scores <sup>23</sup>
Investigator-graded hirsutism scores (women only) <sup>24</sup>
Investigator-graded striae§
Strength
Sit-to-stand test (lower extremity function) <sup>25,26</sup>
Hand-grip strength test <sup>27</sup>
Psychiatric health/cognitive function
Beck depression inventory <sup>28</sup>
Trail making test (cognitive function) <sup>29,30</sup>
Quality of life
SF-36 health survey (quality of life) <sup>31</sup>

\*Data for glucose and insulin levels at time 0, and 30, 60, 90, 120 min and total AUC.

†Haemoglobin A1c at baseline and Weeks 16 and 24.

‡Cushingoid appearance (facial plethora, moon-shaped facies, dorsal cervical fat pad) is graded on a 0–3 scale: 0 = absent, 1 = mild, 2 = moderate, 3 = severe.

§Striae graded for each of 5 body sites (abdomen, thighs, low back, axillae/arms, chest/breast) on a 0–4 scale:

0 = absent.

1 = mild with up to 2 striae, lightly violaceous or red, and up to 1 cm in width.

2 = moderate with 3 or 4 striae, violaceous or red, and >1 cm in width.

3 = major/widespread with 5 or more striae, deeply violaceous, and >1 cm in width.

4 = severe/extensive striae that become confluent with marked skin atrophy.

bone mineral density were available for the baseline and final visits only.

### Statistics

As previously described,<sup>8</sup> the data review board evaluated, in a blinded fashion, the data from the clinical parameters shown in Table 1. An *a priori* positive response rate  $\geq 30\%$  was used to test the overall rate of response to mifepristone; if the lower bound of the one-sided 95% confidence interval exceeded this threshold, the end-point was considered statistically significant. Patients with at least 30 days of treatment [ $n = 46$ , modified intention-to-treat (mITT) population] were evaluated in analyses of efficacy as

previously described.<sup>8</sup> Photographs of 34 consenting patients were available for review. A termination visit was conducted for all patients completing the study as well as for patients with an early termination (ET) date (Week 24/ET,  $n = 40$ ). Data were imputed using last observation carried forward (LOCF) where indicated. *Post hoc* analyses were conducted to assess the time course of GCR using the Cochran–Mantel–Haenszel (CMH) chi-square test controlled for nonindependence and multiple ratings within patients, and adjusted by the number of degrees of freedom. The impact of age and gender on response was examined by repeated measures analysis mixed effects model and the effect of radiation using Fisher's exact test. We examined the contribution of predictive factors on GCR in univariate and multivariate models. Factors with meaningful correlations ( $P \leq 0.1$ ) were used to build a multivariate regression analysis model. The agreement of raters for visits was evaluated by intraclass correlation coefficient (ICC); the  $P$  value was assigned using Fisher transformation for the ICC. The ICC provides an estimate of the difference between how much agreement was actually observed compared with that expected by chance alone. ICC values range between  $-1$  and  $1$  ( $<0$  less than chance agreement;  $0.01$ – $0.2$  slight;  $0.21$ – $0.40$  fair;  $0.41$ – $0.60$  moderate;  $0.61$ – $0.80$  substantial;  $0.81$ – $1.00$  almost perfect

agreement).<sup>9,10</sup> Other statistical testing used paired  $t$ -tests, Wilcoxon signed rank tests and chi-square tests where appropriate. Statistical testing was conducted using SAS statistical software (version 9.2; Cary, NC, USA) and Microsoft Excel 2010 (Redmond, WA, USA).

## Results

### Patients

The study population consisted of 46 patients (mITT population,  $n = 32$  female and 14 male), including 34 patients who completed the study (i.e. through to Week 24 of treatment).<sup>8</sup> The baseline characteristics of patients are shown in Table 2. There were 40 patients with Cushing's disease who had failed prior transphenoidal surgery, including 17 who received prior radiation therapy ( $34.4 \pm 26.2$  months [mean  $\pm$  SD] prior to enrolment). There were three patients with ectopic ACTH syndrome and three with adrenal carcinoma as the cause of CS. Twenty-five of the 46 patients had T2DM and/or impaired fasting glucose (C-DM) while 21 had a diagnosis of hypertension without T2DM and/or IGT.<sup>8</sup> Of the 25 C-DM patients, 19

**Table 2.** Baseline characteristics of patients in the SEISMIC (Study of the Efficacy and Safety of Mifepristone in the Treatment of Endogenous Cushing's Syndrome) study. ( $n = 46$ )

Sex, $n$ (%)		Race, $n$ (%)		Ethnicity, $n$ (%)	
Male	14 (30.4)	African American	6 (13.0)	Hispanic	4 (8.7)
Female	32 (69.6)	White	40 (87.0)	Not Hispanic	42 (91.3)
Age, years		Height, cm		Weight, kg	
Mean (SD)	44.5 (11.7)	Mean (SD)	167 (11.0)	Mean (SD)	99.5 (29.9)
Median	44.5	Median	166	Median	92.4
Minimum; maximum	26; 71	Minimum; maximum	143.5; 190.5	Minimum; maximum	61.3; 198.7
Body mass index, kg/m <sup>2</sup>		Waist circumference, cm		ACTH (pm)*	
Mean (SD)	35.8 (10.1)	Mean (SD)	119 (20.7)	Mean (SD)	14.9 (13.9)
Median	33.5	Median	115	Median	11.0
Minimum; maximum	24.1; 66.4	Minimum; maximum	88.5; 178.4	Minimum; maximum	1.5; 75.9
24-h UFC (nmol/24 h)		Serum cortisol (nm)		Late-night salivary cortisol (nm)	
Mean (SD)	939.4 (2 166.9)	Mean (SD)	651.1 (275.9)	Mean (SD)	13.8 (24.8)
Median	327.2	Median	598.7	Median	8.3
Minimum; maximum	118.3; 13 156.3	Minimum; maximum	251.1; 1680.2	Minimum; maximum	2.8; 140.7
Clinical signs and symptoms, $n$ (%)					
Cushingoid appearance	45 (97.8)	Striae, hirsutism, acne	27 (58.7)		
Proximal muscle weakness	25 (54.3)	Low bone mass†	12 (26.1)		
Psychiatric symptoms	23 (50)				

\* $N = 43$  for ACTH; adrenal carcinoma patients ( $n = 3$ ) excluded from calculation. To convert values of ACTH to pg/ml, divide by 0.22. To convert urinary free cortisol (UFC) to  $\mu\text{g}/24$  h, divide by 2.759. To convert cortisol to  $\mu\text{g}/\text{l}$ , divide by 27.59.

†T below  $-1.0$  at any vertebral site.

had coexisting hypertension.<sup>8</sup> At the start of the study, 15 of 25 C-DM patients were receiving an antidiabetic medication, and 30 of the 40 patients with a diagnosis of hypertension were taking antihypertensive medication.

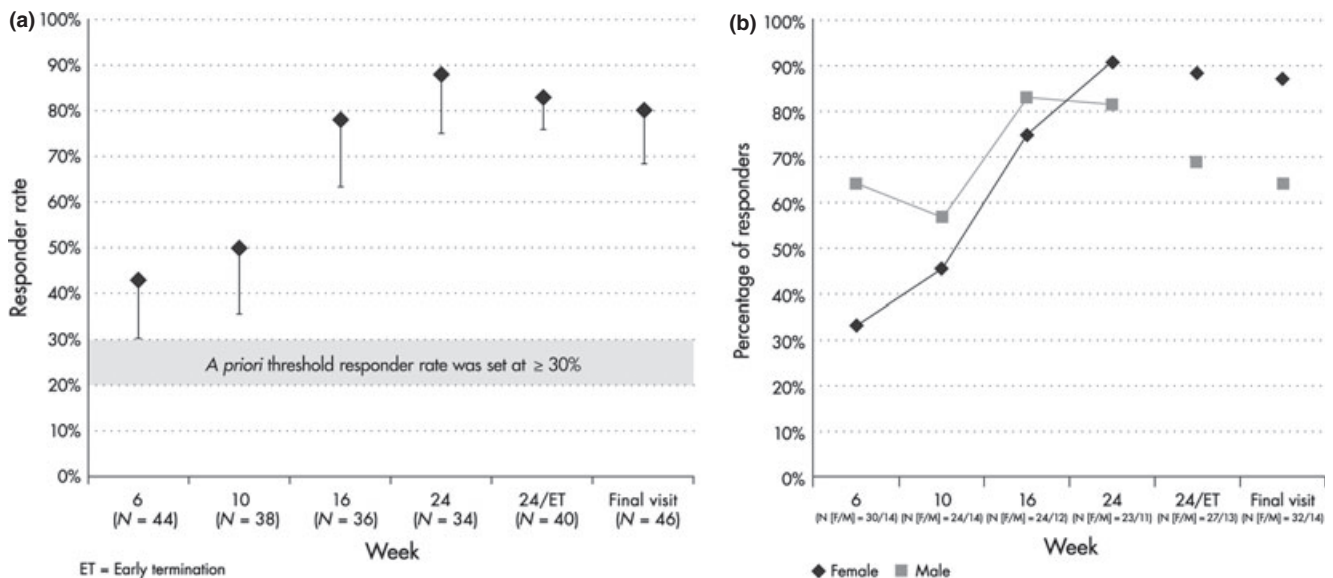
### Global clinical response

The proportion of patients with a positive GCR increased over time with the greatest percentage of patients having a positive GCR at Week 24 (88%,  $P < 0.001$  for trend, CMH chi-square; Fig. 1). The prevalence of positive GCR was 83% ( $n = 40$ ) at the study termination visit (Week 24/ET) and was 80% ( $n = 46$ ) at the last evaluated visit (LOCF). Gender was associated with the time to first positive GCR as well as the improvement in the response over time (Fig. 1). As a group, the women tended to have a slower onset of positive GCR compared with men ( $P = 0.02$  for first response by gender and  $P = 0.003$  for gender effect over time, repeated measures analysis). The prevalence of positive GCR at the study termination visit (Week 24/ET) and the last evaluated visit was numerically higher in women compared with men but the differences were not statistically significant (Fig. 1). At baseline, we observed no differences in the clinical severity of disease between men and women based upon the frequency of typical signs and symptoms of CS; the median number of signs and symptoms was four in both men and women (male:female: Cushingoid appearance 100%:97%,  $P = 0.5$ ; increased body weight 93%:100%,  $P = 0.13$ ; proximal muscle weakness 71%:47%,  $P = 0.12$ ; low bone mass 29%:25%,  $P = 0.8$ ; psychiatric symptoms 64%:44%,  $P = 0.2$ ; hirsutism and/or pigmented striae and/or acne 43%:66%,  $P = 0.15$ , chi-square). Men tended to have higher pretreatment urinary free

cortisol (UFC) levels but the differences were not statistically different from women (UFC in men [mean  $\pm$  SD, median]: 1815  $\pm$  3,664, 436 nmol/24 h vs UFC in women 552  $\pm$  830, 265 nmol/24 h;  $P = 0.22$ , Wilcoxon). Removal from the analysis of two men with very high baseline UFC levels due to ectopic ACTH syndrome resulted in UFC levels (mean  $\pm$  SD, median: 472  $\pm$  345, 339 nmol/24 h) similar to those in women ( $P = 0.66$ , Wilcoxon). At the study termination visit (Week 24/ET) and the last evaluated visit, the prevalence of GCR was similar among CD patients with and without prior pituitary radiation (Week 24/ET: 86.7% vs 77.3%,  $P = 0.68$ , Fisher's exact; last evaluated visit: 88.2% vs 73.9%,  $P = 0.43$ , Fisher's exact, respectively). GCR was not significantly associated with cohort (C-DM vs C-HT;  $P = 0.26$ , repeated measures analysis) or age ( $P = 0.96$ , repeated measures analysis), and there was insufficient statistical power to assess the effects of race or ethnicity (Table 2).

### Rater concordance

Overall, there was 100% agreement among the three raters in 121 of 158 (76.6%) evaluated patient visits, and there was a similar degree of full agreement among raters at each assessed visit (Week 6, 79.5%; Week 10, 71.1%; Week 16, 72.2%; Week 24/ET, 82.5%). The ICC for rater agreement overall was 0.652 ( $P < 0.001$ , Fisher transformation), indicating a substantial degree of overall agreement. Discordance occurred in 37 reviewed visits of which there was only a single case of complete discordance (individual scores of +1, 0, -1; median 0). In the remaining 36, two of three raters agreed and the median scores were +1 in 19 cases, 0 in 16 cases and -1 in one case.



**Fig. 1** Global clinical response rate over time and by gender. (a) The prevalence of patients scored as responders increased over time ( $P < 0.001$ ). Week 24/ET represents all patients who completed the study or had an early termination visit. Final visit indicates the last visit evaluated (last observation carried forward). The *a priori* threshold responder rate for significance was set at  $\geq 30\%$ . Error bars represent the lower end of the one-sided 95% confidence interval. (b) The global clinical response (GCR) by gender through Week 24 differed by gender ( $P = 0.02$  for first GCR, and  $P = 0.003$  for effect over time). Week 24/ET represents all patients who completed or had an early termination visit. Final visit indicates the last visit evaluated (last observation carried forward).

### Correlates of GCR ratings

Table 3 shows the factors that predicted GCR scores in univariate analyses. Entering these variables into a multiple logistic regression model led to only four significant predictors of GCR response: change in baseline body weight ( $P < 0.0001$ ), change in diastolic blood pressure ( $P < 0.0001$ ), change in 120-min glucose on an oral glucose tolerance test (oGTT) ( $P = 0.0003$ ) and change in Cushingoid appearance ( $P = 0.022$ ). Reduction in antidiabetic medication was only modestly associated with GCR scores in univariate analyses ( $P = 0.0631$ ) and was not a significant predictor in the final model. Reduction in antihypertensive medication was not associated with GCR in the univariate analyses ( $P = 0.124$  C-HT cohort and  $0.4197$  C-DM cohort). When gender was entered into the model as an interacting term, the results of the logistic regression were the same.

Cushingoid appearance was evaluated by investigator-graded assessments as well as by photographs available for the 34 of the 46 patients who consented to standardized photography. At

**Table 3.** Predictors of positive global clinical response (GCR)

Factor*	Cohort†	Univariate <i>P</i>	Multivariate <i>P</i>
Body weight change from baseline	Combined	<0.0001	<0.0001
Diastolic blood pressure (BP) change from baseline	C-HT	0.0001	
Diastolic BP change from baseline	Combined	0.0001	<0.0001
Systolic BP change from baseline	Combined	0.0002	
Cushingoid appearance change from baseline	Combined	0.0004	0.022
120-min glucose change from baseline	Combined	0.0004	0.0003
Systolic BP change from baseline	C-HT	0.0011	
Systolic BP at visit	C-HT	0.0049	
Systolic BP at visit	Combined	0.0242	
Fasting glucose change from baseline	Combined	0.0254	
Diastolic BP at visit	C-HT	0.0271	
Waist change from baseline	Combined	0.0318	
SF-36 Physical Composite Score change from baseline	Combined	0.0339	
Diastolic BP at visit	Combined	0.0466	
Diabetes medication reduction at visit	C-DM	0.0631	
Diabetes medication reduction at visit	C-DM w/HTN	0.0855	
Diastolic BP change from baseline	C-DM w/HTN	0.0954	

\*Only factors with  $P \leq 0.1$  in univariate analysis are shown.

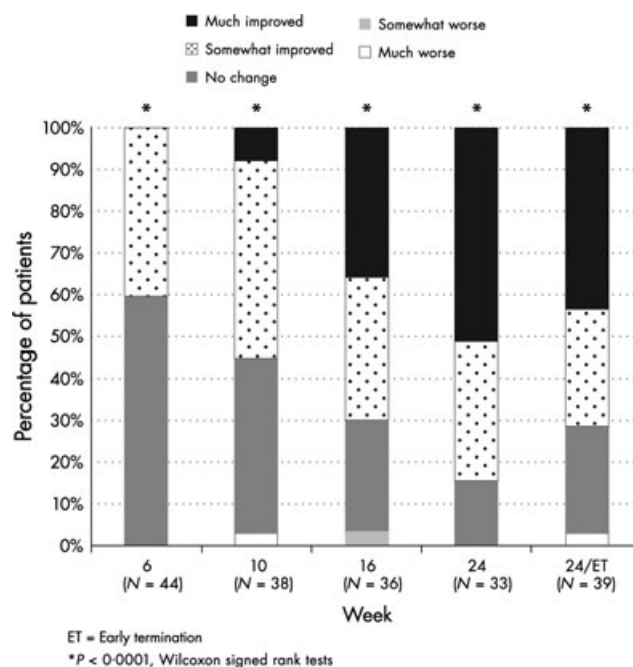
†C-DM = patients with diabetes mellitus/impaired glucose tolerance; C-HT = patients with hypertension alone; Combined = C-DM and C-HT; C-DM w/HTN = C-DM patients with a diagnosis of hypertension.

baseline, 17.4% of patients had severe Cushingoid features, 50% had substantial Cushingoid features, 30.4% had mild Cushingoid features and 2.2% had absent Cushingoid features. As shown in Fig. 2, there was a progressive improvement in Cushingoid appearance, as assessed by the clinical site investigators ( $P < 0.0001$ , Wilcoxon). Improvements in body weight, blood pressure and glucose have been previously reported.<sup>8</sup> Although quality of life was not a significant correlate of positive GCR in the multiple logistic regression model, there was a univariate association with the Physical Composite Score (Table 3). Evaluation of the SF-36 subdomains demonstrated improvement in seven of eight measures (Fig. 3).

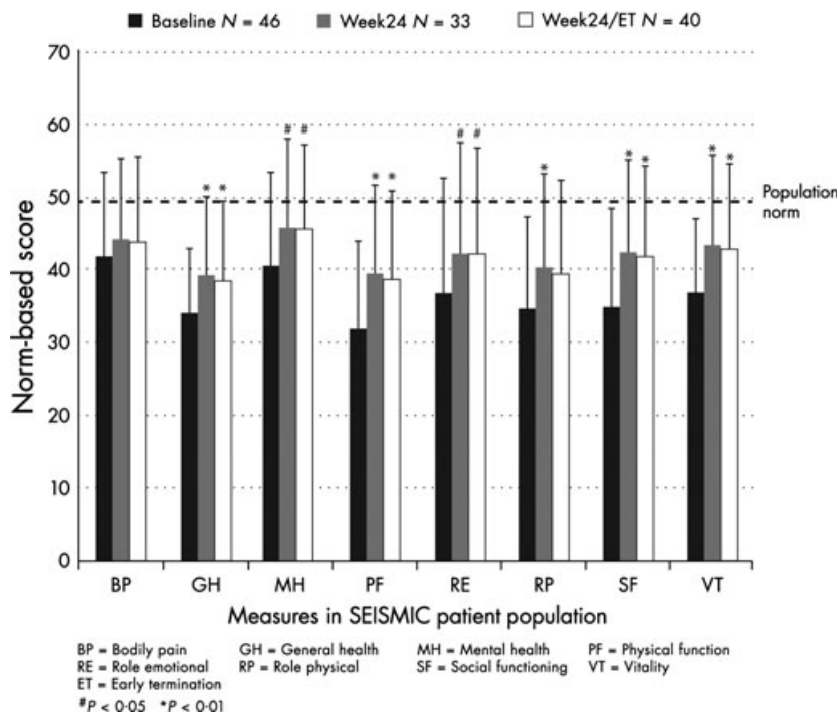
### Discussion

The hypercortisolism of CS results in a multisystem disorder with serious clinical sequelae. Medical therapy is usually relegated to an adjuvant role following incomplete surgical extirpation or recurrence of the tumoural source of hormonal excess; the drugs employed have typically consisted of agents that lower cortisol secretion and have variable efficacy and tolerability.<sup>4,11,12</sup> Mifepristone has a distinct mechanism of action: by antagonizing the effect of cortisol at the receptor level, its clinical effectiveness is not dependent on cortisol reduction. Mifepristone has been recently reported to be highly effective in improving hyperglycaemia associated with CS and many of the clinical manifestations of the disorder<sup>8</sup> and thus was recently approved for use in the United States.

In the recently reported SEISMIC study, 87% of patients with CS achieved the key secondary study end-point of improvement



**Fig. 2** Clinical improvement over time. Progressive improvement in Cushingoid appearance in the patient population as assessed by the clinical site investigators. *P*-values are compared to baseline.



**Fig. 3** Quality of life measures (SF-36). Clinical improvement over time based on evaluation of the SF-36 subdomains demonstrated improvement in 7 of 8 measures in the patient population. *P*-values are compared to baseline.

in GCR during treatment with mifepristone.<sup>8</sup> The benefits of medical treatment in CS should be viewed in the context of a chronic disease that may be stable or may progress. Although the statistical assessment of efficacy used in the primary analyses of the study assessed GCR improvement at any visit, 37% of patients had positive GCR by Week 10 that persisted through study end, whereas only 6.5% of patients had a positive GCR during the study that was not maintained<sup>8</sup>. The additional analyses of the clinical response data reported here demonstrate a progressive clinical benefit of mifepristone, with the greatest proportion of responses by the end of the study at 24 weeks. While twice as many women were evaluated than men, the data suggest that there may be a gender-dependent effect with mifepristone as there was a more rapid clinical benefit in men, although there was no gender difference by the end of the study. Males have been reported to have a more severe expression of CS,<sup>13</sup> potentially leading to a different response to glucocorticoid blockade; however, we did not detect such differences in the baseline severity of CS in our study population. Nonetheless, gender-specific differences in the effectiveness of drugs are well described although poorly understood,<sup>14</sup> and our findings suggest that a greater length of therapy may be required to document the maximal clinical benefit for women. Although mifepristone produces rapid and robust improvements in dysglycaemia in CS patients,<sup>8,15</sup> we did not find that the presence of T2DM/IGT influenced overall GCR. Because of the flexible dose escalation scheme used in the study and the time lag inherent in the reversal of the biological effects of hypercortisolism, we are unable to parse out the independent effects of dose and time on clinical improvement. The most common dose associated with first global clinical response was 600 mg, but 30% of patients required higher doses.<sup>8</sup>

Reduction in the use of antidiabetic and/or antihypertensive medications is an important factor in gauging clinical improvement in CS. At baseline in this study, 15 patients were taking antidiabetic medications,<sup>8</sup> 30 were taking antihypertensive medications and nine were taking both. Fourteen patients required spironolactone to manage hypokalaemia during mifepristone treatment.<sup>16</sup> Lack of statistical power and possible confounding from use of spironolactone may explain why reductions in antihypertensive medications were not a significant predictor of the GCR in the multiple regression model. High UFC levels at baseline (>500 nmol/24 h) increased the likelihood of requiring treatment for hypokalaemia, and the degree of increase in UFC during mifepristone was inversely correlated with serum potassium<sup>16</sup>.

This study used an independent physician rater board to review and assess clinical response with the goal to have a key study end-point of clinical assessment that simulated the clinical practice setting of mifepristone use in CS. We noted full agreement between the clinical raters in over three-quarters of their evaluations. The statistical assessment of concordance (ICC) among raters demonstrated a substantial degree of agreement, which compares well with other clinical rating systems.<sup>17,18</sup> Sonino *et al.*<sup>19</sup> evaluated a clinical rating tool for disease severity in CS that was based on a three-point ordinal assessment of each of eight clinical domains, six of which were similar to the ones used in our system (fat/weight, skin/appearance, muscle weakness, mood, hypertension, diabetes), and a commonly used global assessment tool that employed a nine-point scale for overall illness severity and change.<sup>20</sup> With two raters assessing 14 patients before and after the cure of CS, there was an ICC of 0.95 and 0.87, respectively, for the eight-domain tool and lower values for the global scales (0.86 for severity and 0.67 for

change). Our rating system is not directly comparable to that of Sonino<sup>19</sup> because we used three rather than two raters, and because Sonino's raters directly evaluated the patients, whereas our system used raters that did not evaluate patients in person. Nonetheless, our system with independent raters performed similarly to the global system used by Sonino,<sup>19</sup> demonstrating the practicality of our system in a clinical trial setting.

The GCR assessments in our study required raters to provide an overall clinical appraisal of response after considering numerous laboratory and clinical data. In multiple regression analyses, we noted four significant predictors of a graded positive GCR (changes in weight, 120-min oGTT glucose, diastolic blood pressure and investigator-graded Cushingoid appearance). It is interesting that two of these (glucose and weight) were also two independent end-points that demonstrated the most substantial quantitative improvement in the SEISMIC study (i.e. weight and glucose).<sup>8</sup> Although photographs were not available for all rated patients, we believe that investigator-graded Cushingoid appearance was a significant predictor of GCR because it was complementary to the photographs that were reviewed. It is not possible to assess a correlation between the qualitative photographs and the quantitative appearance ratings. Nonetheless, the change in appearance was an important element in the raters' judgment of change in clinical status. Gender did not appear to modify the predictors of GCR although it was associated with the pace of response suggesting that there were no differential treatment effects between men and women. Although there were improvements in the quality-of-life measures at the end of the study, average quality-of-life scores remained below the population norm and did not appear to influence significantly the raters' assessment of overall response to treatment. Longer treatment duration may result in further improvement. Persistent impaired quality of life is common in CS even after long-term biochemical remission postsuccessful surgery.<sup>21,22</sup> Our findings suggest that the assessment of multiple clinical variables can be used by clinicians to assess mifepristone response and dosing in CS.

There are several limitations to our analysis. This *Post hoc* analysis is subject to bias because it was not a predefined analysis. Although these data should be interpreted with caution, they highlight important performance characteristics of a major study end-point. The methodology used to determine GCR was chosen to integrate multiple clinical variables into a single rating to provide a clinically relevant assessment of response to therapy (e.g. to simulate clinical judgement in a practice situation). This introduces a degree of subjectivity as some end-points reviewed were semiquantitative or qualitative (e.g. photographs) and it also reduces statistical power. The open-label design of the study would potentially make the reviewers more likely to assess a positive GCR at any particular evaluated visit. We believe that such bias was limited because the reviewers were blind to the visit sequence (i.e. duration of treatment and dose) when patient data were reviewed and clear patient improvement was shown over time when analysed in chronological sequence. Our use of a stepped approach for the logistic regression method (univariate followed by multivariate) may have led us to miss an interacting variable in the final model; however, other approaches

were not appropriate for the small sample size and large number of variables in our data set.

In conclusion, treatment of Cushing's syndrome with mifepristone results in early and progressive clinical improvement in most patients. Strong drivers of clinical response included changes in weight, glucose, blood pressure and physical appearance. Male gender may be associated with a faster response to treatment. The independent review of variables assessing the clinical response demonstrated robust performance as an end-point in the clinical trial and provides evidence that appraisal of numerous clinical parameters can be utilized when managing a CS patient on mifepristone.

## Acknowledgements

Laurence Katznelson and David E. Schteingart received research grants from Corcept Therapeutics for their role as investigators. D. Lynn Loriaux, Glenn D. Braunstein, and David Feldman were consultants to Corcept Therapeutics. Coleman Gross is an employee of Corcept Therapeutics. This study was sponsored by Corcept Therapeutics.

## References

- 1 Shimon, I., Ram, Z., Cohen, Z.R. *et al.* (2002) Transsphenoidal surgery for Cushing's disease: endocrinological follow-up monitoring of 82 patients. *Neurosurgery*, **51**, 57–61.
- 2 Hammer, G.D., Tyrrell, J.B., Lamborn, K.R. *et al.* (2004) Transsphenoidal microsurgery for Cushing's disease: initial outcome and long-term results. *Journal of Clinical Endocrinology and Metabolism*, **89**, 6348–6357.
- 3 Atkinson, A.B., Kennedy, A., Wiggam, M.I. *et al.* (2005) Long-term remission rates after pituitary surgery for Cushing's disease: the need for long-term surveillance. *Clinical Endocrinology*, **63**, 549–559.
- 4 Tritos, N.A. & Biller, B.M.K. (2012) Advances in medical therapies for Cushing's syndrome. *Discovery Medicine*, **13**, 171–179.
- 5 Fleseriu, M. & Petersen, M. (2012) Medical management of Cushing's disease: what is in the future? *Pituitary*, **15**, 330–341.
- 6 Castinetti, F., Morange, I., Jaquet, P. *et al.* (2008) Ketoconazole revisited: a preoperative or postoperative treatment in Cushing's disease. *European Journal of Endocrinology*, **158**, 91–99.
- 7 Colao, A., Petersenn, S., Newell-Price, J. *et al.* (2012) A 12-month phase 3 study of pasireotide in Cushing's disease. *The New England Journal of Medicine*, **366**, 914–924.
- 8 Fleseriu, M., Biller, B.M., Findling, J.W. *et al.* (2012) Mifepristone, a glucocorticoid receptor antagonist, produces clinical and metabolic benefits in patients with Cushing's syndrome. *Journal of Clinical Endocrinology and Metabolism*, **97**, 2039–2049.
- 9 Landis, R.J. & Koch, G.G. (1977) The measurement of observer agreement for categorical data. *Biometrics*, **33**, 159–174.
- 10 Fleiss, J.L. & Cohen, J. (1973) The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational Psychological Measurement*, **33**, 613–619.
- 11 van der Pas, R., de Herder, W.W., Hofland, L.J. *et al.* (2012) New developments in the medical treatment of Cushing's syndrome. *Endocrine-Related Cancer*, **19**, R205–R223.

- 12 Feelders, R.A. & Hofland, L.J. (2013) Medical treatment of Cushing's disease. *Journal of Clinical Endocrinology and Metabolism*, **98**, 425–438.
- 13 Giraldi, F.P., Moro, M. & Cavagnini, F. (2003) Gender-related differences in the presentation and course of Cushing's disease. *Journal of Clinical Endocrinology and Metabolism*, **88**, 1554–1558.
- 14 Franconi, F., Brunelleschi, S., Steardo, L. *et al.* (2007) Gender differences in drug responses. *Pharmaceutical Research*, **55**, 81–95.
- 15 Wallia, A., Colleran, K., Purnell, J.Q. *et al.* (2013) Improvement in insulin sensitivity during mifepristone treatment of Cushing syndrome: early and late effects. *Diabetes Care*, **36**, e147–e148.
- 16 Auchus, R., McPhaul, M.J., Findling, J.W. *et al.* (2012) Predicting and managing hypertension and hypokalemia in Cushing syndrome during mifepristone therapy. *Endocrine Reviews*, **33** (03 Meeting abstracts), MON–450.
- 17 Slora, E.J., Bocian, A.B., Herman-Giddens, H. *et al.* (2009). Assessing inter-rater reliability (IRR) of Tanner staging and orchidometer use with boys: a study from PROS. *Journal of Pediatric Endocrinology & Metabolism*, **22**, 291–299.
- 18 Lucky, A.W., Barber, B.L., Girman, C.J. *et al.* (1996) A multirater validation study to assess the reliability of acne lesion counting. *Journal of the American Academy of Dermatology*, **35**, 559–565.
- 19 Sonino, N., Boscaro, M., Fallo, F. *et al.* (2000) A clinical index for rating severity in Cushing's syndrome. *Psychotherapy and Psychosomatics*, **69**, 216–220.
- 20 Kellner, R. (1972) Improvement criteria in drug trials with neurotic patients. *Psychological Medicine*, **2**, 73–80.
- 21 Lindsay, J.R., Nansel, T., Baid, S. *et al.* (2006) Long-term impaired quality of life in Cushing's syndrome despite initial improvement after surgical remission. *Journal of Clinical Endocrinology and Metabolism*, **91**, 447–453.
- 22 van Aken, M.O., Pereira, A.M., Biermasz, N.R. *et al.* (2005) Quality of life in patients after long-term biochemical cure of Cushing's disease. *Journal of Clinical Endocrinology and Metabolism*, **90**, 3279–3286.
- 23 Doshi, A., Zaheer, A. & Stiller, M.J. (1997) A comparison of current acne grading systems and proposal of a novel system. *International Journal of Dermatology*, **36**, 416–418.
- 24 Hatch, R., Rosenfield, R.L., Kim, M.H. *et al.* (1981) Hirsutism: implications, etiology and management. *American Journal of Obstetrics and Gynecology*, **140**, 815–830.
- 25 Butler, A.A., Menant, J.C., Tiedemann, A.C. *et al.* (2009) Age and gender differences in seven tests of functional mobility. *Journal of Neuroengineering and Rehabilitation*, **6**, 31.
- 26 Lord, S.R., Murray, S.M., Chapman, K. *et al.* (2002) Speed, balance and psychological status in older people. *Journal of Gerontology*, **57A**, M539–M543.
- 27 Mathiowetz, V., Kashman, N., Volland, G. *et al.* (1985) Grip and pinch strength: normative data for adults. *Archives of Physical Medicine and Rehabilitation*, **66**, 69–74.
- 28 Beck, A.T., Steer, R.A. & Brown, G. (1996) Manual for the Beck Depression Inventory-II. Psychological Corporation, San Antonio, TX.
- 29 Reitan, R.M. & Wolfson, D. (1985) Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation. Neuropsychological Press, Tuscon, AZ.
- 30 Tombaugh, T.N. (2004) Trail Making Test A and B: normative data stratified by age and education. *Archives of Clinical Neuropsychology*, **19**, 203–214.
- 31 Ware, J.E. Jr, Kosinski, M., Bjorner, J.B. *et al.* (2007) User's manual for the SF-36v2™ Health Survey, 2nd edn. Quality Metric Incorporated, Lincoln, RI.